

DOI:10.19651/j.cnki.emt.2210828

基于改进生成对抗网络的无监督晶圆缺陷检测

李 阳 蒋三新

(上海电力大学电子与信息工程学院 上海 201306)

摘要: 为实现晶圆表面缺陷的无监督检测,提出了一种改进生成对抗网络的无监督晶圆表面缺陷检测模型,该模型通过目标图像与重构图像之间的差异来检测缺陷。该方法使用带有两层跳跃连接和记忆模块的编码器-解码器卷积神经网络来搭建生成器,跳跃连接用以捕获多尺度的输入图像特征,记忆模块对潜在特征实施约束,扩大真实缺陷样本与重构样本间的距离。该方法还通过改进判别器网络结构,使模型轻量化。实验结果表明,该模型能够准确分辨具有缺陷的晶圆样本,ROC曲线下的面积值达到0.934,与已有的无监督学习检测方法相比性能更优,同时判别器网络的参数量和计算量分别降低到1 M和60 M以下。

关键词: 生成对抗网络;无监督学习;缺陷检测;晶圆;记忆模块

中图分类号: TP391.41;TN407 **文献标识码:** A **国家标准学科分类代码:** 510.4050

Unsupervised wafer defect detection based on improved generative adversarial network

Li Yang Jiang Sanxin

(School of Electronic and Information Engineering, Shanghai University of Electric Power, Shanghai 201306, China)

Abstract: In order to realize unsupervised detection of wafer surface defects, an unsupervised wafer surface defect detection model with improved generative adversarial network was proposed. The model detected the defects by the difference between the target image and the reconstructed image. In this method, an encoder-decoder convolutional neural network with two layers of skip connections and memory module was used to build the generator. The skip connections were used to capture multi-scale input image features, and the memory module was used to constrain latent characteristics to enlarge the distance between real defect samples and reconstructed samples. The method also makes the model lightweight by improving the discriminator network structure. Experimental results show that the proposed model can accurately distinguish the defective wafer samples, and the area value under the ROC curve reaches 0.934, which is better than the existing unsupervised learning detection methods, and the parameters and flops of the discriminator network is reduced to less than 1 M and 60 M.

Keywords: generative adversarial network; unsupervised learning; defect detection; wafer; memory module

0 引 言

近年来,随着半导体行业的飞速发展^[1],芯片产量逐年增加,芯片生产中的良品率至关重要。晶圆作为芯片的重要材料,在加工制作的过程中,经过熔融、切割、研磨、刻蚀等众多复杂的工艺流程,流程中的任何异常都会导致晶圆缺陷的产生^[2]。为确保晶圆表面的质量,缺陷检测^[3]工序显得尤为重要。

现今的半导体制造工厂多采用人工借助显微镜的方式来检测晶圆缺陷,然而,由于晶圆数量多且体积微小,这种检测方式不仅需要大量的人力物力,而且效率低、检测不准

确。传统的缺陷检测算法大都基于机器视觉和监督学习。基于机器视觉的缺陷检测,往往采用常规图像处理算法或人工设计特征加分类器的方式^[4],由于缺陷的多样性,该方法很难对缺陷特征进行完整的建模和迁移,复用性不大。基于监督学习的缺陷检测,常使用带有标签的缺陷数据去训练分类器^[5],该方法依赖大量的缺陷样本以及标注数据,特别在工业生产中,缺陷样本出现概率极低,并需要专业人员对其标注,制作成本高。因此,无监督的缺陷检测显得尤为重要。

无监督的缺陷检测是深度学习领域非常重要的一项研究内容。生成对抗网络(generative adversarial network,

收稿日期:2022-07-26

GAN)和生成对抗训练框架^[6]已经成功地应用于对复杂和高维分布的现实数据进行建模。GAN 的这一特性表明它们可以成功地应用于无监督的缺陷检测。目前基于 GAN 的无监督缺陷检测的任务是使用对抗训练过程建模正常行为,并通过测量异常得分来检测缺陷。异常是指数据中不符合定义良好的正常行为概念的模式^[7],晶圆表面缺陷数据也属于异常数据中的一种。

用 GAN 做无监督缺陷检测的模型常采用基于重构的方法,重构的方法常依赖于一个假设,即仅在正常数据上学习的模型不能准确地表示和重构缺陷^[8]。Schlegl 等^[9]首次将标准的 GAN 用于检测异常,生成器学习潜在特征到现实样本的映射,使生成器学习真实样本的分布。判别器用以区分真实样本和生成样本,通过对抗训练,使生成的重构图像更接近真实样本的分布。但是该模型每对一个新的图像进行检测时,为了生成一个与输入图像尽可能相似的正常图像,都需要通过比较重构图像与输入图像的差异去更新潜在特征的参数。为了解决不断更新参数的问题,Zenati 等^[10]引入 BiGAN^[11]的网络架构,通过添加编码器的方法增加反向映射学习,将数据映射回潜在表示,通过图像空间到潜在空间的映射学习来捕获正常样本的分布。Akçay 等^[12]采用编码器-解码器-编码器来搭建生成器网络,以捕获正常图像在潜在空间的分布。同年,Akçay 等^[13]又提出一种带有跳跃连接的编解码器卷积神经网络,通过对抗训练使模型学习生成高质量的正常图像。另外基于 GAN 的检测模型中,判别器的结构常采用 DCGAN^[14]的判别器结构,DCGAN 的判别器使用普通卷积,模型的参数数量和计算量大。

为此,本文提出了一种改进生成对抗网络的无监督学习方法来检测晶圆表面的缺陷。本方法仅对正常晶圆图像进行训练,以学习正常样本的特征分布,并通过目标图像与重构图像之间的差异来检测缺陷。模型包含生成器和判别器两个网络,生成器采用编码器-解码器的卷积神经网络结构,并将编码器的最后两层与解码器的前端两层特征跳跃连接,以捕获多尺度的输入图像特征,使得重构图像包含更多输入图像的细节。并且在生成器的潜在空间添加记忆模块,对潜在特征实施约束,以记录正常样本的相关信息,扩大真实缺陷样本与重构样本间的距离,提高缺陷的重构误差。另外,通过改进判别器网络结构,减少了模型的参数数量和计算量,使网络更加轻量化。实验表明,本文提出的模型在晶圆表面缺陷检测任务中性能优越。

1 检测方法

1.1 总体模型

本文提出了一种用于检测晶圆表面缺陷的无监督模型,如图 1 所示,该模型由生成器 G 和判别器 D 两个网络组成,生成器用于图像重构,判别器用于区分真实图像与重构图像,并用来衡量输入图像与重构图像在潜在空间的差异。生成器由 3 个部分组成:编码器 G_e (用于提取输入图像特征并生成潜在表示)、记忆模块(具有记忆和寻址操作)和解码器 G_d (用于生成图像)。给定一个输入图像 x ,编码器首先对输入图像进行编码,获得的潜在特征 z 作为查询元,记忆模块通过寻址操作查询与 z 最相关的项输出,得到新的潜在特征 \hat{z} ,最后将 \hat{z} 送进解码器网络中进行上采样,将上采样得到的前端两层特征与编码器最后两层特征进行

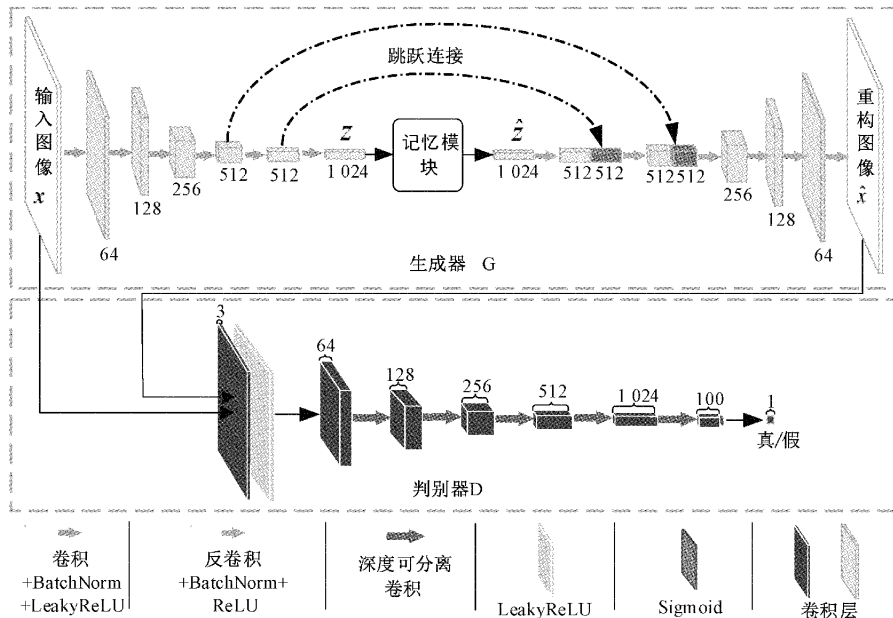


图 1 模型整体框架

融合后解码输出重构图像 \hat{x} 。判别器获取输入图像与重构图像,通过卷积下采样获得它们的潜在表示,并预测其类别标签。通过对抗训练,优化生成器与判别器,最小化重构误差。同时更新记忆模块用以记录正常样本的相关信息。给定一个测试样本,记忆模块仅能检索到与正常样本最相关的信息输入解码器网络中进行重构。因此,重构后的样本分布接近正常样本,使得正常样本的重构图像与原始图像差异较小,缺陷样本的重构差异较大,这将作为检测缺陷的依据。

1.2 编码器与解码器

编码器网络 G_e 用于提取输入图像的特征信息,输入图像 x 在编码器网络中通过 6 个区域,其中每个区域包括卷积层、BatchNorm 层和 LeakyReLU 激活函数,经过下采样将高维的输入图像 x 映射为低维的潜在特征 z , z 是输入图像的特征表示。定义 X 为真实图像空间, Z 为潜在空间,为了让编码器网络捕获输入数据的分布,训练编码器将输入图像 $x \in X$ 映射到潜在空间 Z ,使 $G_e: x \rightarrow z, z \in Z$ 。

解码器网络 G_d 用于将潜在特征逆映射为重构图像。与编码器网络对称,解码器网络也包含 6 个区域,其中每个区域包括反卷积层、BatchNorm 层以及 ReLU 激活函数,通过上采样将低维的潜在特征 \hat{z} 逆映射为高维的重构图像 \hat{x} 。另外,如图 1 所示,将编码器网络的最后两层与解码器网络的前端两层特征跳跃连接^[15],通过这种连接方式,实现了层与层之间信息的传递,保留了多尺度的信息,从而获得更好的重构效果。

1.3 记忆模块

为了记录输入正常样本的相关信息,在生成器的潜在空间添加记忆模块,潜在特征 z 经记忆模块查询得到新的特征 \hat{z} ,将 \hat{z} 作为解码器网络的输入来重建样本。本文参考文献[16]提出的寻址方法,将 z 到 \hat{z} 的过程定义为寻址过程,将 z 和 \hat{z} 定义为查询元。记忆模块的结构如图 2 所示。

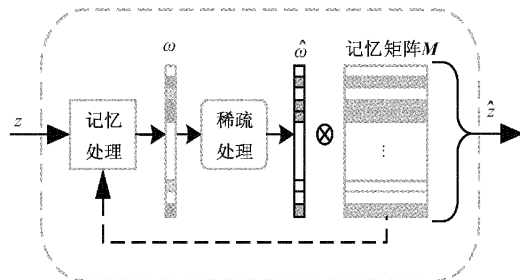


图 2 记忆模块

定义向量空间为 \mathbb{R} ,图 2 中记忆矩阵 $M \in \mathbb{R}^{N \times C}$, N 代表 M 矩阵包含 N 个行向量,每个行向量表示一个记忆元素,维度为 C ,该维度等于经编码器输出的潜在特征 z 的维度。记忆处理模块计算每一个查询元 z 与记忆元素

的相似性得到注意力系数 ω ,计算公式如下:

$$\omega_i = \text{Softmax}(d(z, m_i)) \quad (1)$$

$$d(z, m_i) = \frac{z m_i^T}{\|z\| \cdot \|m_i\|} \quad (2)$$

其中, $z \in \mathbb{Z}^{1 \times C}, m_i \in \mathbb{R}^{1 \times C}$ 。 $d(z, m_i)$ 表示查询元 z 与记忆元素 m_i 之间的余弦相似度, ω_i 表示 ω 的第 i 个元素, m_i 表示 M 矩阵的第 i 个行向量。训练时只输入正常样本,使记忆矩阵 M 仅记录正常样本的分布。

由于一些小而密集的注意力系数 ω 仍会使缺陷样本的潜在特征 z 产生较好的重构效果,为了避免这些参数对重构效果的影响,通过稀疏处理的方式去掉这些参数,计算公式如下:

$$\hat{\omega}_i = \begin{cases} \omega_i, & \omega_i > \zeta \\ 0, & \text{其他} \end{cases} \quad (3)$$

训练时需要将注意力参数 ω 进行梯度更新,为了便于梯度求解,利用 ReLU 函数将式(3)改写为一个连续的形式,计算公式如下:

$$\hat{\omega}_i = \frac{\text{ReLU}(\omega_i - \zeta) \cdot \omega_i}{|\omega_i - \zeta| + \delta} \quad (4)$$

其中, $\zeta = 0.0025, \delta = 10^{-12}$ 。将稀疏处理之后得到的注意力系数 $\hat{\omega}$ 与记忆矩阵相乘得到新的潜在特征 \hat{z} 。其式如(5)所示。

$$\hat{z} = \hat{\omega} M = \sum_i \hat{\omega}_i m_i \quad (5)$$

训练时由于记忆矩阵 M 只记录了正常样本的分布,测试时将缺陷样本输入网络,经编码器网络编码获得潜在特征 z ,再通过记忆模块的寻址计算,获得的新特征 \hat{z} 仅仅包含正常样本分布的最相关信息输入解码器网络进行重构,这样就能实现对缺陷样本有较大的重构误差,进而提高缺陷检测效果。

1.4 判别器

模型中的第 2 个网络如图 1 所示为判别器 D ,在训练过程中,判别器的任务是试图区分真实样本与生成样本,并进行分类。另外,判别器除了作为分类器外,还作为特征提取器,提取输入图像与重构图像的特征,并获得它们的潜在特征表示。本文借鉴文献[17]提出的深度可分离卷积思想,改进了判别器网络结构,与方法^[9,12-15]使用的普通卷积结构对比如图 3 所示。改进后的判别器使用卷积结构如图 3(右)所示,从图中可以看出,与普通卷积不同,该卷积是一种可分解的卷积操作,其分解为两个更小的卷积操作:深度卷积(dw)和逐点卷积(pw)。深度卷积采用 4×4 的卷积核,一个卷积核只负责一个通道。逐点卷积采用 1×1 的卷积核,每个卷积操作都包括卷积层、BatchNorm 层和 LeakyReLU 激活函数。

判别器的具体网络结构如表 1 所示。首先是一个 4×4 的普通卷积,后面是堆叠的深度卷积和逐点卷积,输入特征经过步长为 2 的深度卷积进行下采样,再通过步长

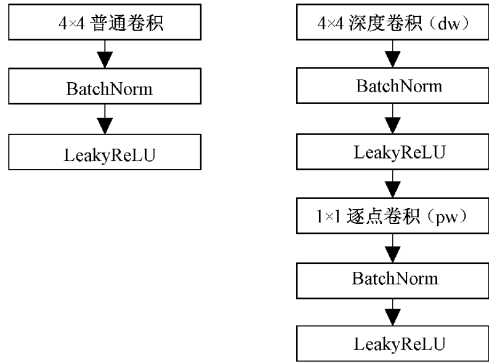


图 3 普通卷积与改进后卷积结构对比

为 1 的逐点卷积改变特征维度,在最后一个深度卷积层,采用步长为 1 的 4×4 卷积核将特征维度变为 1×1。根据预测类别大小加上 3×3 普通卷积层,最后通过 Sigmoid 激活函数输出判别结果。

表 1 判别器网络结构

卷积方式	步长	滤波器形状	输入大小
Conv	2	4×4×3×64	128×128×3
Conv dw	2	4×4×64×64	64×64×64
Conv pw	1	1×1×64×128	32×32×64
Conv dw	2	4×4×128×128	32×32×128
Conv pw	1	1×1×128×256	16×16×128
Conv dw	2	4×4×256×256	16×16×256
Conv pw	1	1×1×256×512	8×8×256
Conv dw	2	4×4×512×512	8×8×512
Conv pw	1	1×1×512×1 024	4×4×512
Conv dw	1	4×4×1 024×1 024	4×4×1 024
Conv pw	1	1×1×1 024×100	1×1×1 024
Conv	1	3×3×100×1	1×1×100
Sigmoid	1	Classifier	1×1×1

1.5 损失函数

本模型仅在正常样本上进行训练,并在正常和缺陷样本上进行测试。其目的是希望模型通过学习正常样本分布,在图像和潜在空间中能正确地重构出正常样本。相反,模型未对缺陷样本进行训练,记忆模块只能寻址得到正常样本的最相关信息,因此模型只能以正常样本特征来重建缺陷部分。基于上述分析,模型对缺陷样本的重构图像 \hat{x} 及其经判别器得到的潜在特征表示 $D_{100}(\hat{x})$ 具有较高的重建损失。为此,本文设置了对抗损失、内容损失、潜在损失以及注意力损失,并将其加权求和,每个损失值在整个训练过程中都有各自的贡献。图 4 为网络总体损失训练流程。

假设给定一个包含 K 个样本的数据集 $\{x_i\}_{i=1}^K$, \hat{x} 表示与每个训练样本 x 对应的重构样本。为保证模型对正

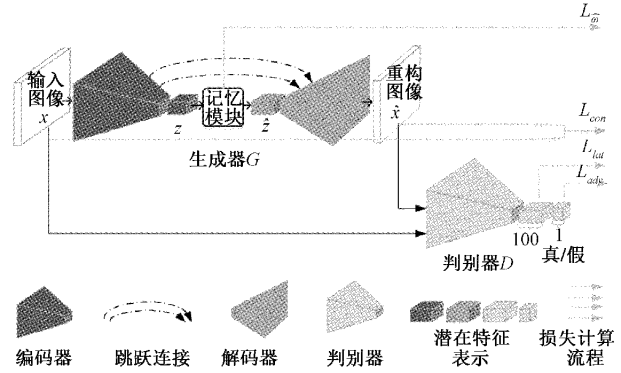


图 4 网络总体损失训练流程

常样本的重建能力,本文参考文献[6]中提出的对抗损失来训练模型。该损失保证了生成器 G 尽可能重建出真实的正常图像,而判别器 D 对输入的真实样本 x 与重构样本(假样本) \hat{x} 进行分类,其任务是通过最小化生成器 G 和最大化判别器 D 来达到 $\min_G \max_D L_{adv}$ 目的,对抗损失公式如下:

$$L_{adv} = \frac{1}{K} \sum_{i=1}^K [\log D(x_i) + \log(1 - D(\hat{x}_i))] \quad (6)$$

对抗损失要求模型生成更逼真的真实图像,但没能学习到输入图像的内容。为了使生成器充分捕捉输入正常样本的数据分布,保证输入与重构图像内容上相似,对每个输入图像和重构图像用 L_1 范数归一化,内容损失公式如下:

$$L_{con} = \frac{1}{K} \sum_{i=1}^K \|x_i - \hat{x}_i\|_1 \quad (7)$$

通过对抗损失以及内容损失,使模型能够生成内容相似且更逼真的真实图像。除此之外,为了能更全面地了解输入样本分布,使输入图像 x 和重构图像 \hat{x} 在潜在空间尽可能相似,提取 x 和 \hat{x} 经判别器 D 下采样得到的维度为 100 的特征,其特征表示分别为 $D_{100}(x)$ 和 $D_{100}(\hat{x})$,并使用范数 L_2 归一化这两个特征表示,潜在损失公式如下:

$$L_{lat} = \frac{1}{K} \sum_{i=1}^K \|D_{100}(x_i) - D_{100}(\hat{x}_i)\|_2 \quad (8)$$

为了降低模型对缺陷样本的重建能力,加强记忆模块对正常样本相关信息的记录,除了采用式(4)对注意力系数 ω 进行处理外,还使用约束信息熵对稀疏处理后的注意力系数 $\hat{\omega}$ 进行归一化处理,进一步约束注意力系数 ω 的稀疏性,注意力损失公式如下:

$$L_{atn} = \frac{1}{K} \sum_{i=1}^K -\hat{\omega}_i \log(\hat{\omega}_i) \quad (9)$$

综上,模型训练的总损失函数是上述 4 个部分的加权总和,计算公式如下:

$$L = \lambda_{adv} L_{adv} + \lambda_{con} L_{con} + \lambda_{lat} L_{lat} + \alpha L_{atn} \quad (10)$$

式中: λ_{adv} 、 λ_{con} 和 λ_{lat} 作为加权参数,用于调整各部分损失函数对总损失的影响, α 是训练中的超参数,通过多次实验

调参得到 $\lambda_{adv} = 1, \lambda_{con} = 50, \lambda_{lat} = 1, \alpha = 2 \times 10^{-1}$ 是最理想的结果。训练时,通过反向传播和梯度下降优化来更新记忆矩阵 \mathbf{M} 。反向传播的过程中,只有当每个注意力系数 ω_i 非零时,每个记忆元素 m_i 才是非零的。

1.6 异常分数

为了在测试和后续的检测任务中发现有缺陷的晶圆样本,本文使用文献[18]中提出的异常评分。对于给定的测试图像 x' ,异常分数由重构分数 $R(x')$ 和潜在表征分数 $L(x')$ 组成,其计算公式如下:

$$R(x') = L_{con}(x') \quad (11)$$

$$L(x') = L_{lat}(x') \quad (12)$$

其中, $R(x')$ 是根据式(7)计算得出,用以衡量输入图像与重构图像在内容上的相似程度。 $L(x')$ 是根据式(8)计算得出,用以衡量输入图像与重构图像在潜在空间的差异。最终,异常分数 $S_A(x')$ 的计算公式如下:

$$S_A(x') = \lambda_1 R(x') + \lambda_2 L(x') \quad (13)$$

其中, λ 是控制分数相对重要的加权参数。测试时计算测试集 D_{test} 中每个测试样本 x' 的异常得分,并表示为异

常分数集合 $S_A = \{S_{A_i}; S_A(x'_i), x'_i \in D_{test}\}$, S_{A_i} 表示第 i 个异常分数值, x'_i 表示测试集 D_{test} 中第 i 个待测样本, $S_A(x'_i)$ 表示第 i 个测试样本的异常得分。最后,使用特征缩放,将异常分数 S_A 中每个元素缩放到 $[0, 1]$ 之间。因此,对单个测试样本的异常得分如式(14)所示。

$$\tilde{S}_A = \frac{S_A(x') - \min(S_A)}{\max(S_A) - \min(S_A)} \quad (14)$$

式中: $\max(S_A)$ 和 $\min(S_A)$ 分别为异常得分集合 S_A 中的最大分数值和最小分数值,最终以 \tilde{S}_A 来对验证集和测试集进行评估。

2 实验结果及分析

2.1 数据集

实验采用晶圆数据集 5 263 张,将正常晶圆图像作为正样本,测试集由 427 个正样本以及 427 个负样本组成。缺陷晶圆各类别部分图像如图 5 所示,数据集包含外来异物、异常电镀、原料不良、药液残留、裸芯造成的 7 种不同大小的缺陷,具有一定的多样性。

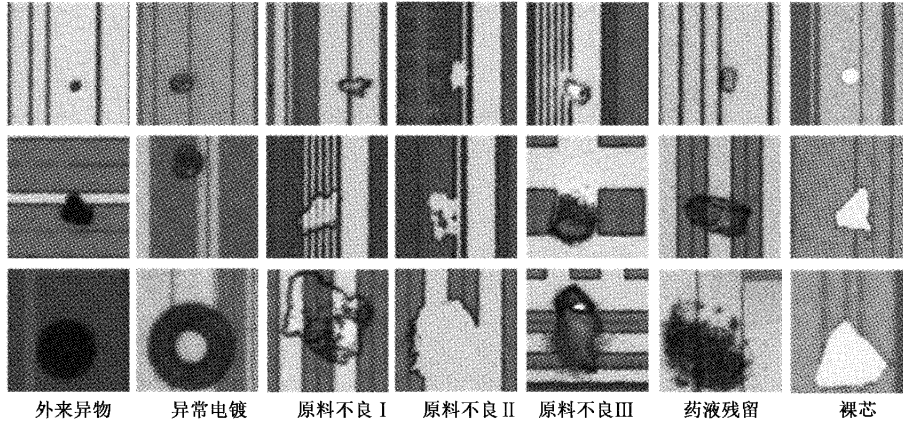


图 5 晶圆缺陷数据示例

为了进一步评估本模型的性能,还对公共数据集 Mvtec AD^[19] 中的物体类缺陷进行了实验。Mvtec AD 数据集包含来自不同领域的 10 种物体的高分辨率彩色图像,包含用于训练的正常图像和用于测试的异常图像。异常包含划痕、污染、部件变形等不同类型的缺陷。

2.2 实验平台及评估指标

实验所使用的平台及运行环境配置如表 2 所示。在模型测试阶段,模型性能通过精确率 (precision)、召回率 (recall)、 F_1 分数以及计算受试者工作特征曲线 (receiver operating characteristics curve, ROC)^[20] 下的面积 (area under curve, AUC) 来进行评估。精确率、召回率、 F_1 分数定义分别如下:

$$R_{precision} = \frac{N_{TP}}{N_{TP} + N_{FP}} \quad (15)$$

$$R_{recall} = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad (16)$$

$$R_{F_1} = \frac{2 \times R_{precision} \times R_{recall}}{R_{precision} + R_{recall}} \quad (17)$$

式中: N_{TP} 表示正常样本分类正确次数, N_{FP} 表示缺陷样本分类错误次数, N_{FN} 表示正常样本分类错误次数。ROC 曲线是根据不同阈值下的真阳性率 (true positive rate, TPR) 和假阳性率 (false positive rate, FPR) 绘制,其反映了分类器对正例和负例的分类能力。另外,在测试样本不均衡的情况下,AUC 依然能对分类器做出合理的评估。

表 2 实验平台及运行环境配置

项目	配置
操作系统	Ubuntu 20.04
CPU	Intel Xeon Silver 4210
GPU	NVIDIA GeForce RTX 2080 Ti
RAM	32 GB
深度学习框架	Pytorch 1.3.1
编译语言	Python 3.7

2.3 训练细节及参数设置

训练模型输入图片大小均为 128×128 像素, 批量大小为 8, 使用 Adam 优化器对式 (10) 中的训练目标 L 进行优化, 初始学习率为 0.000 2, 并伴有 λ 衰减, 动量 $\beta_1 = 0.5$ 、 $\beta_2 = 0.999$ 。通过多次实验调参得到式 (13) 中异常分数 $S_A(x')$ 中加权参数 $\lambda_1 = 0.9$ 、 $\lambda_2 = 0.1$ 。

2.4 实验结果分析

为了对本文所提出的算法进行评估, 利用所构建的晶圆数据集与传统基于 GAN 的无监督检测模型 AnoGAN^[9]、EGBAD^[10]、GANomaly^[12]、Skip-GANomaly^[13] 进行对比实验, 所得 ROC 曲线如图 6 所示, 对比结果如表 3 所示。

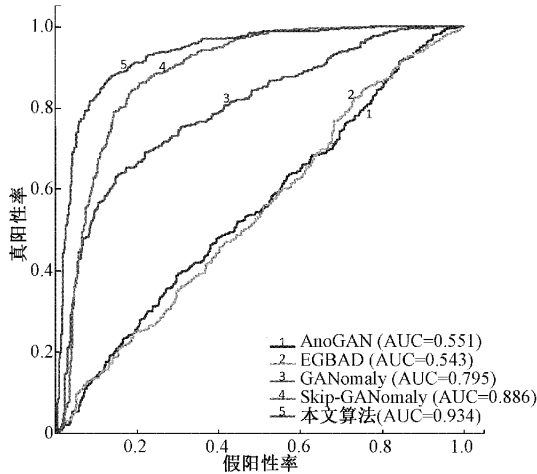


图 6 各模型的 ROC 曲线

表 3 各模型在不同评估指标下的性能对比

模型	精确率	召回率	F_1	AUC
AnoGAN ^[9]	0.514	0.526	0.520	0.551
EGBAD ^[10]	0.505	0.511	0.508	0.543
GANomaly ^[12]	0.711	0.722	0.716	0.795
Skip-GANomaly ^[13]	0.808	0.827	0.817	0.886
本方法	0.866	0.870	0.868	0.934

由图 6 可以看出, 本文所提方法优于现有模型, 对于存在缺陷的晶圆图像, 本文所提模型实现了更高的 AUC 值, 相比 Skip-GANomaly 模型, 本模型的 AUC 提升了 4.8%。

表 3 从精确率 (precision)、召回率 (recall)、 F_1 分数以及 AUC 这 4 个指标来全方位评估本模型性能。从表 2 可以得出本模型相比其他模型性能更优。AnoGAN 和 EGBAD 模型不具备对较强数据的重建能力且定义的异常分数可解释性差。GANomaly 模型学习能力差, 对正常样本的重建能力弱, 导致正常样本的异常分数值较高, 正常样本被误分类的概率高。Skip-GANomaly 模型可以正确捕获图像的分布, 但对缺陷样本分布同样具有较好的重建

能力, 导致缺陷样本重构误差小, 检测效果不佳。

表 4 所示为本模型与 1.4 节提到的 3 种方法的判别器参数量与计算量结果对比, 结果表明本模型改进判别器网络结构, 减少了模型的参数量与计算量, 使模型更加轻量化。

表 4 各模型判别器网络参数量与计算量对比

模型	参数量/M	计算量/M
AnoGAN ^[9]	11.16	550.45
GANomaly ^[12]	11.16	550.45
Skip-GANomaly ^[13]	12.79	552.08
本文方法	0.92	50.93

图 7 所示为测试数据中正常样本与缺陷样本的异常分数直方图以及拟合的曲线。从图中可以看出本模型在正、负样例的异常分数上产生了明显的分离。且最终得到的建议阈值为 0.17, 当异常分数大于 0.17 时, 认为晶圆图像中存在缺陷, 反之认为晶圆正常。

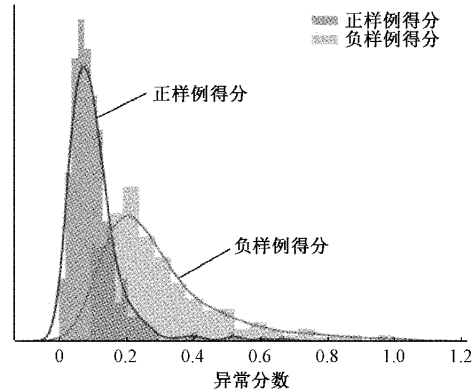


图 7 正常和缺陷样例的得分直方图

表 5 所示为各模型对不同大小缺陷进行检测的 AUC 值。从表中可以观察到相比其他模型, 本模型对不同大小缺陷的检测性能更优。但为了保证模型的重构效果, 模型使用了两层跳跃连接线来捕捉图像特征, 因此当输入目标图像出现较小缺陷时, 重构图像与目标图像之间的差异较小, 所以小目标缺陷的检测效果比中等和大目标缺陷的检测效果略弱。

表 5 不同大小缺陷检测的 AUC 值

模型	小缺陷	中等缺陷	大缺陷
AnoGAN ^[9]	0.473	0.572	0.633
EGBAD ^[10]	0.498	0.550	0.560
GANomaly ^[12]	0.623	0.829	0.924
Skip-GANomaly ^[13]	0.725	0.899	0.924
本方法	0.850	0.957	0.975

图 8、9 所示为实验中晶圆数据集的示例测试原始图像、重构图像以及对应的原始图像与重构图像之间的像

素级差异。由于本文在生成器的最后两层使用跳跃连接来提取图像多尺度信息,因此能够在验证与测试时对正常样例与缺陷样例进行重构。计算原始图像与重构图像之间的差异,由于本模型在生成器潜在空间添加记忆模

块来记录正常样本相关信息,因此从图中可以观察到模型会根据学习到的正常晶圆图像的特征信息来对目标图像进行重建,从而在缺陷部分会与原始图像有较明显的差异。

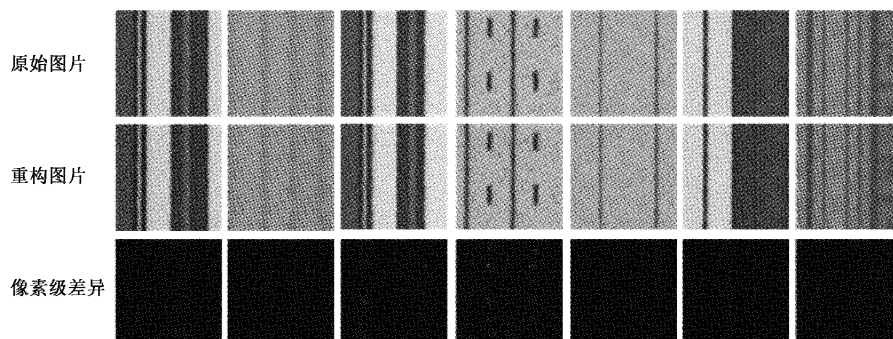


图 8 正常样例检测结果图

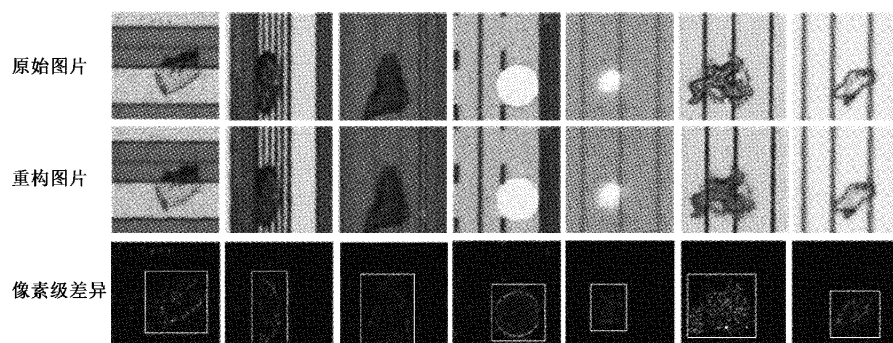


图 9 缺陷样例检测结果图

本文还对公共数据集 Mvtec AD 中的物体类缺陷进行了检测,AUC 结果如表 6 所示。由表 6 可以看出,除了物

体类 Cable、Pill 和 Toothbrush 外,所提模型的结果优于之前的工作。

表 6 在 Mvtec AD 数据集上物体类缺陷检测的 AUC 结果

类别	AnoGAN ^[9]	EGBAD ^[10]	GANomaly ^[12]	Skip-GANomaly ^[13]	本方法
Bottle	0.690	0.683	0.892	0.806	0.921
Cable	0.530	0.558	0.757	0.707	0.744
Capsule	0.580	0.578	0.732	0.738	0.745
Hazelnut	0.500	0.761	0.785	0.876	0.899
Metal Nut	0.500	0.527	0.700	0.719	0.742
Pill	0.620	0.669	0.743	0.811	0.777
Screw	0.350	0.420	0.746	0.911	0.937
Toothbrush	0.570	0.680	0.653	0.964	0.950
Transistor	0.670	0.557	0.792	0.776	0.819
Zipper	0.590	0.645	0.745	0.743	0.834

2.5 消融实验

为了验证生成器模型跳跃连接的数量对检测结果的影响,本文分别设置了 0、1、2、3、4、5 个跳跃连接线,在晶圆数据集上进行消融实验,原始图像与重构图像之间的像素级差异结果如图 10 所示。

由图 10 可以看到,随着跳跃连接数量的增加,模型的重构能力增强,输入样本的重构效果越来越好,原始图像与重构图像之间的差异越来越小。连接数为 0、1 层时,模型对正常样本的重建能力较弱,正常样本的重构差异较大;连接数量为 3、4、5 层时,模型对缺陷样本中的缺陷部分

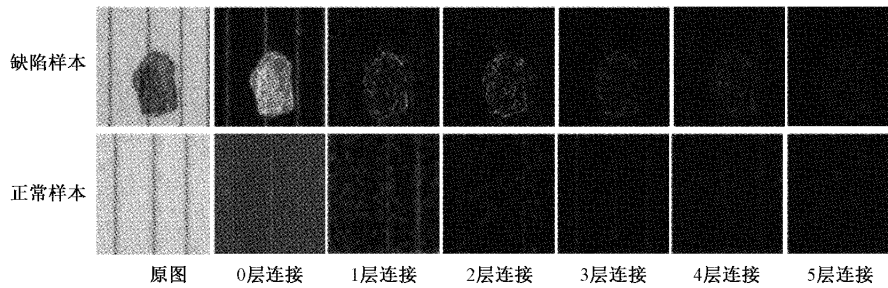


图 10 不同跳跃连接数的重构差异图

重建能力变强,使得缺陷样本的重构差异较小,导致模型检测缺陷的能力较弱。模型的检测性能用 AUC 来评估,如表 7 所示。因此,考虑到模型的性能与检测效果,本文最终选择在生成器网络上添加两层跳跃连接。

表 7 不同跳跃连接数下模型的 AUC 值

跳跃数	AUC
0	0.603
1	0.910
2	0.934
3	0.870
4	0.827
5	0.784

为了验证在模型中添加记忆模块以及对判别器网络进行改进后对结果的影响,在晶圆数据集上进行消融实验,AUC 结果如表 8 所示。从中可以看出,在生成器的潜在空间添加记忆模块,对潜在特征实施约束,使模型记录正常样本相关信息,有效提高了模型的检测性能。另外,改进后的判别器对最终的实验结果影响较小,但是却大大减小了模型的参数量与计算量,如表 4 所示。因此,综合考虑下搭建了本文的检测模型。

表 8 记忆模块和改进后的判别器在模型中的影响

记忆模块	改进后的判别器	AUC
		0.829
✓		0.938
	✓	0.826
✓	✓	0.934

3 结 论

在生成对抗网络(GAN)的算法框架上,针对晶圆在芯片加工制作过程中的缺陷检测的问题,以及传统检测方法与已有的无监督学习的检测方法存在的局限性,提出了一种改进生成对抗网络的无监督检测模型。模型仅依靠正常晶圆样本进行训练,使模型学习正常样本分布,测试时将待测晶圆输入模型,获得对应的重构图像。计算待测图

像与重构图像之间的差异来实现对晶圆缺陷的检测。通过实验分析,本文所提出的方法对晶圆缺陷检测实现了 0.934 的 AUC 值,相较已有的无监督学习检测方法平均提升 24%,表明所提出的方法在性能上优于以往无监督检测模型,为晶圆生产时的验片工艺提供了一种易于工程实践的自动化检测方案,并对后续将无监督学习方法应用于晶圆的疵点检测等任务具有指导意义。目前模型难以应用于需要缺陷精确位置的场景中,后续将进一步研究利用重构图像实现对缺陷位置精确定位。

参考文献

- [1] 于志斌, 胡泓. 基于 YOLO 算法与机器视觉的晶圆片表面缺陷检测研究[J]. 新型工业化, 2021, 11(12): 114-117.
- [2] 吴斌, 卢笑蕾, 余建波. 晶圆表面缺陷模式的在线探测与自适应识别研究[J]. 计算机工程与应用, 2016, 52(17): 261-265, 270.
- [3] 王延年, 刘宏涛, 刘航宇, 等. 基于改进 U-Net 的太阳能电池图像缺陷检测方法[J]. 电子测量技术, 2021, 44(14): 117-121.
- [4] JIN Q, CHEN L. A survey of surface defect detection of industrial products based on a small number of labeled data[J]. ArXiv Preprint, 2022, ArXiv:2203.05733.
- [5] 陶显, 侯伟, 徐德. 基于深度学习的表面缺陷检测方法综述[J]. 自动化学报, 2021, 47(5): 1017-1034.
- [6] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks [J]. ArXiv Preprint, 2014, ArXiv:1406.2661.
- [7] DI MATTIA F, GALEONE P, DE SIMONI M, et al. A survey on gans for anomaly detection [J]. ArXiv Preprint, 2019, ArXiv:1906.11632.
- [8] ZONG B, SONG Q, MIN M R, et al. Deep autoencoding gaussian mixture model for unsupervised anomaly detection [C]. International Conference on Learning Representations, 2018.
- [9] SCHLEGL T, SEEBÖCK P, WALDSTEIN S M, et al. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery [C]. International Conference on Information Processing in

- Medical Imaging. Springer, Cham, 2017: 146-157.
- [10] ZENATI H, FOO C S, LECOAT B, et al. Efficient gan-based anomaly detection [J]. ArXiv Preprint, 2018, ArXiv:1802.06222.
- [11] DONAHUE J, KRÄHENBÜHL P, DARRELL T. Adversarial feature learning[J]. ArXiv Preprint, 2016, ArXiv:1605.09782.
- [12] AKCAY S, ATAPOUR-ABARGHOUEI A, BRECKON T P. Ganomaly: Semi-supervised anomaly detection via adversarial training[C]. Asian Conference on Computer Vision. Springer, Cham, 2018: 622-637.
- [13] AKCAY S, ATAPOUR-ABARGHOUEI A, BRECKON T P. Skip-ganomaly: Skip connected and adversarially trained encoder-decoder anomaly detection[C]. 2019 International Joint Conference on Neural Networks(IJCNN), IEEE, 2019: 1-8.
- [14] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks [J]. ArXiv Preprint, 2015, ArXiv:1511.06434.
- [15] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [16] WESTON J, CHOPRA S, BORDES A. Memory networks[J]. ArXiv Preprint, 2014, ArXiv:1410.3916.
- [17] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications [J]. ArXiv Preprint, 2017, ArXiv:1704.04861.
- [18] BIAN J, HUI X, SUN S, et al. A novel and efficient CVAE-GAN-based approach with informative manifold for semi-supervised anomaly detection[J]. IEEE Access, 2019, 7: 88903-88916.
- [19] BERGMANN P, FAUSER M, SATTLEGGER D, et al. MVTec AD-A comprehensive real-world dataset for unsupervised anomaly detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 9592-9600.
- [20] LING C X, HUANG J, ZHANG H. AUC: A statistically consistent and more discriminating measure than accuracy [C]. International Joint Conference on Artificial Intelligence, 2003.

作者简介

李阳, 硕士研究生, 主要研究方向为深度学习与图像处理、缺陷检测。

E-mail: liyang0920@mail.shiep.edu.cn

蒋三新(通信作者), 博士, 讲师, 主要研究方向为机器视觉、深度增强学习等。

E-mail: samjoe_2018@shiep.edu.cn